ELL2, a New Member of an ELL Family of RNA Polymerase II Elongation Factors

This application claims the benefit of the filing date of provisional application 60/038,447 filed on February 19, 1997, which is herein incorporated by reference.

Field of the Invention

This invention relates to newly identified polynucleotides, polypeptides encoded by them and to the use of such polynucleotides and polypeptides, and to their production. More particularly, the polynucleotides and polypeptides of the present invention relate to a new member of an ELL family of RNA polymerase II elongation factors, hereinafter referred to as ELL2. The invention also relates to inhibiting or activating the action of such polynucleotides and polypeptides.

Background of the Invention

The elongation stage of eukaryotic messenger RNA synthesis is a major site for the regulation of gene expression (Reines, D. *et al., Trends. Biochem. Sci. 21*:351-355 (1996), Bentley, D.L., *Curr. Opin. Genet. Dev. 5*:210-216 (1995)). Moreover, a growing body of evidence suggests that mis-regulation of elongation may be a key element in a variety of human diseases (Aso, T. *et al., J. Clin. Invest. 97*:1561-1569 (1996)).

To date, one virally encoded protein (Tat) and five cellular proteins (SII, P-TEFb, TFIIF, Elongin (SIII), and ELL) have been defined biochemically and shown to be capable of controlling the activity of the RNA polymerase II elongation complex. Among these elongation factors, three have been implicated in human disease. The HIV-1 encoded Tat protein is required for efficient transcription of HIV-1 genes and for productive infection by the virus (Jones, K.A. & Peterlin, B.M., *Annu. Rev. Biochem. 63*:717-743 (1994)). Elongin (SIII) is a potential target for regulation by the product of the von Hippel-Lindau (VHL) tumor suppressor gene, which is mutated in the majority of clear-cell renal

carcinomas and in families with VHL disease, a rare genetic disorder that predisposes individuals to a variety of cancers (Duan, D.R. *et al., Science 269*:1402-1406 (1995), Kibel, A. *et al., Science 269*:1444-1446 (1995)). The *ELL* gene on chromosome 19pl3.1 was originally isolated as a gene that undergoes frequent translocations with the *Drosophila trithorax*-like *MLL* gene on chromosome 11q23 in acute myeloid leukemia (Thirman, M.J. *et al., Proc. Natl. Acad. Sci. U.S.A. 91*:12110-12114 (1994), Mitani, K. *et al., Blood 85*:2017-2024 (1995)).

This indicates that these proteins have an established, proven history as therapeutic targets. Clearly there is a need for identification and characterization of further related proteins which can play a role in preventing, ameliorating or correcting dysfunctions or diseases, including, but not limited to, AIDS and neoplastic disorders, among others.

Summary of the Invention

In one aspect, the invention relates to ELL2 polypeptides and recombinant materials and methods for their production. Another aspect of the invention relates to methods for using such ELL2 polypeptides and polynucleotides. Such uses includes the treatment of neoplastic disorders, among others. In still another aspect, the invention relates to methods to identify agonists and antagonists using the materials provided by the invention, and treating conditions associated with ELL2 imbalance with the identified compounds. Yet another aspect of the invention relates to diagnostic assays for the detection of diseases associated with inappropriate ELL2 activity or levels and mutations in ELL2 that might lead to neoplastic disorders (particularly leukemias).

Brief Description of the Figures

FIG. 1A-1C shows the nucleotide and deduced amino acid sequence of human ELL2 (SEQ ID NOS:1 and 2, repectively).

FIG. 2 shows a comparison of the deduced amino acid sequences of human ELL2 (SEQ ID NO:2) and ELL (SEQ ID NO:7). Identical amino acids are shown in white letters on a black background; similar amino acids (A,S,T,P; D,E,N,Q; H,R,K; I,L,M,V; F,Y,W) are shown in black letters on a grey background.

FIG. 3 shows the localization of the ELL2 elongation activation domain and a summary of ELL2 mutants and their activities in transcription. Wild type ELL2 is diagramed at the bottom of the panel. Conserved regions 1, 2, and 3 (R1, R2, and R3) are indicated by the shaded boxes. The alignment of region 3 with the C-terminal ZO-1 binding domain of occludin (SEQ ID NO:8) was generated with the BESTFIT program of the Genetics Computer Group package, using the symbol comparison table of Gribskov and Burgess (Gribskov, M. & Burgess, R.R., *Nucleic. Acids. Res. 14*:6745-6763 (1986)).

FIG. 4 shows a schematic representation of the pHE4-5 expression vector (SEQ ID NO:33) and the subcloned ELL2 cDNA coding sequence. The locations of the kanamycin resistance marker gene, the ELL2 coding sequence, the oriC sequence, and the *lac*Iq coding sequence are indicated.

FIG. 5 shows the nucleotide sequence of the regulatory elements of the pHE promoter (SEQ ID NO:34). The two *lac* operator sequences, the Shine-Delgarno sequence (S/D), and the terminal *Hind*III and *Nde*I restriction sites (italicized) are indicated.

Detailed Description of the Preferred Embodiments

Definitions

The following definitions are provided to facilitate understanding of certain terms used frequently herein.

"Neoplastic disorder" refers to a disease state which is related to the hyperproliferation of cells. Neoplastic disorders include, but are not limited to, carcinomas, sarcomas and leukemias.

"Protein Activity" or "Biological Activity of the Protein" refers to the metabolic or physiologic function of said ELL2 protein including similar activities or improved activities or these activities with decreased undesirable side-effects. Also included are antigenic and immunogenic activities of said ELL2 protein. Among the physiological or metabolic activities of said protein are the regulation of the activity of the RNA polymerase II elongation complex. As demonstrated in Example 1, ELL2 increases the overall rate of elongation by RNA polymerase II during both promoter-dependent and -independent transcription. Additional activities include the ability to bind components of the RNA polymerase II elongation complex and SH3 domains.

"ELL2 polypeptides" refers to polypeptides with amino acid sequences sufficiently similar to ELL2 protein sequences that they exhibit at least one biological activity of the protein.

"ELL2 gene" refers to a polynucleotide having the nucleotide sequence set forth in SEQ ID NO:1 or the nucleotide sequence encoding the protein as contained in the cDNA insert of ATCC Deposit No. 97863 or allelic variants thereof and/or their complements.

"ELL2 polynucleotides" refers to a polynucleotide containing a nucleotide sequence which encodes an ELL2 polypeptide or fragment thereof or that encodes an ELL2 polypeptide or fragment wherein said nucleotide sequence has at least 95% identity to a nucleotide sequence encoding the polypeptide of SEQ ID NO:2 or the corresponding fragment thereof or which has sufficient identity to a nucleotide sequence contained in SEQ ID NO:1 or contained in the cDNA insert of ATCC Deposit No. 97863 to hybridize under conditions useable for amplification or for use as a probe or marker.

"Antibodies" as used herein includes polyclonal and monoclonal antibodies, chimeric, single chain, and humanized antibodies, as well as Fab fragments, including the products of an Fab or other immunoglobulin expression library.

"Isolated" means altered "by the hand of man" from the natural state. If an "isolated" composition or substance occurs in nature, it has been changed or removed from its original environment, or both. For example, a polynucleotide or a polypeptide naturally present in a living animal is not "isolated," but the same polynucleotide or polypeptide separated from the coexisting materials of its natural state is "isolated", as the term is employed herein. Thus, a polypeptide or polynucleotide produced and/or contained within a recombinant host cell is considered isolated for purposes of the present invention. Also intended as an "isolated polypeptide" or an "isolated polynucleotide" are polypeptides or polynucleotides that have been purified, partially or substantially, from a recombinant host cell or from a native source. For example, a recombinantly produced version of the ELL2 polypeptide can be substantially purified by the one-step method described in Smith and Johnson, *Gene 67*:31-40 (1988).

"Polynucleotide" generally refers to any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. "Polynucleotides" include, without limitation single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, "polynucleotide" refers to triple-stranded regions comprising RNA or DNA or both RNA and DNA. The term polynucleotide also includes DNAs or RNAs containing one or more modified bases and DNAs or RNAs with backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications have been made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically or metabolically modified forms of

polynucleotides as typically found in nature, as well as the chemical forms of DNA and RNA characteristic of viruses and cells. "Polynucleotide" also embraces relatively short polynucleotides, often referred to as oligonucleotides.

"Polypeptide" refers to any peptide or protein comprising two or more amino acids joined to each other by peptide bonds or modified peptide bonds, *i.e.*, peptide isosteres. "Polypeptide" refers to both short chains, commonly referred to as peptides, oligopeptides or oligomers, and to longer chains, generally referred to as proteins. Polypeptides may contain amino acids other than the 20 gene-encoded amino acids. "Polypeptides" include amino acid sequences modified either by natural processes, such as post-translational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched and branched cyclic polypeptides may result from post-translation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as

arginylation, and ubiquitination. See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York, 1993 and Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pgs. 1-12 in POSTTRANSLATTONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, 1983; Seifter *et al.*, "Analysis for protein modifications and nonprotein cofactors", *Methods in Enzymol. 182*:626-646 (1990) and Rattan *et al.*, "Protein Synthesis: Posttranslational Modifications and Aging", *Ann NY Acad Sci 663*:48-62 (1992).

"Variant" as the term is used herein, is a polynucleotide or polypeptide that differs from a reference polynucleotide or polypeptide respectively, but retains essential properties. A typical variant of a polynucleotide differs in nucleotide sequence from another, reference polynucleotide. Changes in the nucleotide sequence of the variant may or may not alter the amino acid sequence of a polypeptide encoded by the reference polynucleotide. Nucleotide changes may result in amino acid substitutions, additions, deletions, fusions and truncations in the polypeptide encoded by the reference sequence, as discussed below. A typical variant of a polypeptide differs in amino acid sequence from another, reference polypeptide. Generally, differences are limited so that the sequences of the reference polypeptide and the variant are closely similar overall and, in many regions, identical. A variant and reference polypeptide may differ in amino acid sequence by one or more substitutions, additions, deletions in any combination. A substituted or inserted amino acid residue may or may not be one encoded by the genetic code. A variant of a polynucleotide or polypeptide may be a naturally occurring such as an allelic variant, or it may be a variant that is not known to occur naturally. Non-naturally occurring variants of polynucleotides and polypeptides may be made by mutagenesis techniques or by direct synthesis.

"Identity" is a measure of the identity of nucleotide sequences or amino acid sequences. In general, the sequences are aligned so that the highest order match is obtained. "Identity" *per se* has an art-recognized meaning and can be

calculated using published techniques. (See, *e.g.*: COMPUTATIONAL MOLECULAR BIOLOGY, Lesk, A.M., ed., Oxford University Press, New York, 1988; BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS, Smith, D.W., ed., Academic Press, New York, 1993; COMPUTER ANALYSIS OF SEQUENCE DATA, PART I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY, von Heinje, G., Academic Press, 1987; and SEQUENCE ANALYSIS PRIMER, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans (Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988)). Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in Guide to Huge Computers, Martin J. Bishop, ed., Academic Press, San Diego, 1994, and Carillo, H. & Lipton, D., *SIAM J Applied Math 48*:1073 (1988). Methods to determine identity and similarity are codified in computer programs. Preferred computer program methods to determine identity and similarity between two sequences include, but are not limited to, GCS program package (Devereux, J., *et al., Nucleic Acids Research 12(i)*:387 (1984)), BLASTP, BLASTN, FASTA (Atschul, S.F., *et al., J Molec Biol 215*:403 (1990)).

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence encoding a ELL2 polypeptide is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the ELL2 polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference

sequence. These mutations of the reference sequence may occur at the 5′ or 3′ terminal positions of the reference nucleotide sequence or anywhere between those terminal positions, interspersed either individually among nucleotides in the reference sequence or in one or more contiguous groups within the reference sequence.

Deposited Material

The invention relates to polypeptides and polynucleotides of a novel ELL2 protein, which is related by amino acid sequence identity to the members of the ELL family of RNA polymerase II elongation factors. The invention relates especially to ELL2 materials having the nucleotide and amino acid sequences set out in SEQ ID NOS:1 and 2, and to the ELL2 nucleotide sequences of the human cDNA in ATCC Deposit No. 97863 and amino acid sequence encoded therein.

A deposit containing a human ELL2 cDNA has been deposited with the American Type Culture Collection, 12301 Park Lawn Drive, Rockville, Maryland 20852, USA, on January 31, 1997, and assigned ATCC Deposit No. 97863. The deposited material is an expression vector referred to as pET-ELL2. This vector was constructed, as described in Example 1, by the insertion of an ELL2 cDNA sequence into the *Sal*I and *Bam*HI sites of M13mpET (Tan, S. *et al.*, *BioTechniques 16*:824-828 (1994)) followed by oligonucleotide-directed mutagenesis (Kunkel, T.A., *Proc. Natl. Acad. Sci. U.S.A. 82*:488-492 (1985)). While the ATCC deposit is believed to contain the ELL2 cDNA sequence shown in SEQ ID NO:1, the nucleotide sequence of the polynucleotide contained in the deposited material, as well as the amino acid sequence of the polypeptide encoded thereby, are controlling in the event of any conflict with any description of sequences herein.

The deposit has been made under the terms of the Budapest Treaty on the international recognition of the deposit of micro-organisms for purposes of patent procedure. The deposit is provided merely as convenience to those of skill in the

art and is not an admission that a deposit is required for enablement, such as that required under 35 U.S.C. § 112.

Polypeptides of the Invention

The ELL2 polypeptides of the present invention include the polypeptide of SEQ ID NO:2, as well as polypeptides and fragments which have activity which have at least 95% identity to the polypeptide of SEQ ID NO:2 or the relevant portion and more preferably at least 96%, 97% or 98% identity to the polypeptide of SEQ ID NO:2 and still more preferably at least 99% identity to the polypeptide of SEQ ID NO:2.

The polypeptides of the present invention are preferably provided in an isolated form.

The polypeptides of the present invention include the polypeptide encoded by the deposited cDNA; a polypeptide comprising amino acids from about 1 to about 640 in SEQ ID NO:2; a polypeptide comprising amino acids from about 2 to about 640 in SEQ ID NO:2; a polypeptide comprising amino acids from about 7 to about 350 in SEQ ID NO:2; a polypeptide comprising amino acids from about 50 to about 389 in SEQ ID NO:2; a polypeptide comprising amino acids from about 443 to about 474 in SEQ ID NO:2; a polypeptide comprising amino acids from about 516 to about 640 in SEQ ID NO:2; as well as polypeptides which are at least 95% identical, and more preferably at least 96%, 97%, 98% or 99% identical to the polypeptides described above and also include portions of such polypeptides with at least 30 amino acids and more preferably at least 50 amino acids.

The ELL2 polypeptides may be a part of a larger protein such as a fusion protein. It is often advantageous to include additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification such as multiple histidine residues, or additional sequence for stability during recombinant production.

Biologically active fragments of the ELL2 polypeptides are also included in the invention. A fragment is a polypeptide having an amino acid sequence that entirely is the same as part but not all of the amino acid sequence of the aforementioned ELL2 polypeptides. As with ELL2 polypeptides, fragments may be "free-standing," or comprised within a larger polypeptide of which they form a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 101-120, 121-140, 141-160, 161-180, 181-200, 201-220, 221-240, 241-260, 261-280, 281-300, 301-320, 321-340, 341-360, 361-380, 381-400, 401-420, 421-440, 441-460, 461-480, 481-500, 501-520, 521-540, 541-560, 561-580, 581-600, 601-620, and 621-640. In this context "about" includes the particularly recited ranges larger or smaller by several, a few, 5, 4, 3, 2 or 1 amino acid at either extreme or at both extremes.

Preferred fragments include, for example, truncation polypeptides having the amino acid sequence of ELL2 polypeptides, except for deletion of a continuous series of residues that includes the amino terminus, or a continuous series of residues that includes the carboxyl terminus or deletion of two continuous series of residues, one including the amino terminus and one including the carboxyl terminus. Also preferred are fragments characterized by structural or functional attributes such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Biologically active fragments are those that mediate protein activity, including those with a similar activity or an improved activity, or with a decreased undesirable activity. Also included are those that are antigenic or immunogenic in an animal, especially in a human.

Thus, the polypeptides of the invention include polypeptides having an amino acid sequence at least 95% identical to that of SEQ ID NO:2 or fragments thereof with at least 95% identity to the corresponding fragment of SEQ ID NO:2 all of which retain the biological activity of the ELL2 protein, including antigenic activity. Included in this group are variants of the defined sequence and fragment. Preferred variants are those that vary from the referents by conservative amino acid substitutions -- *i.e.*, those that substitute a residue with another of like characteristics. Typical such substitutions are among Ala, Val, Leu and Ile; among Ser and Thr; among the acidic residues Asp and Glu; among Asn and Gln; and among the basic residues Lys and Arg, or aromatic residues Phe and Tyr. Particularly preferred are variants in which several, 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination.

The ELL2 polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

Polynucleotides of the Invention

Another aspect of the invention relates to isolated polynucleotides which encode the ELL2 polypeptides and polynucleotides closely related thereto.

ELL2 protein of the invention, is structurally related to other proteins of ELL family of RNA polymerase II elongation factors, as shown by the results of sequencing the cDNA encoding human ELL2 in the deposited clone. The cDNA sequence contains an open reading frame encoding a protein of 640 amino acids with a deduced molecular weight of about 72,354 Da. Thus, ELL2 has a molecular weight of about 72 kilodaltons. ELL2 of SEQ ID NO:2 has about 49% identity over its entirety with ELL.

One polynucleotide of the present invention encoding ELL2 protein may be obtained using standard cloning and screening, from a cDNA library derived

from mRNA in cells of human fetal heart using the expressed sequence tag (EST) analysis (Adams, M.D., *et al., Science* 252:1651-1656 (1991); Adams, M.D., *et al., Nature 355*:632-634; Adams, M.D., *et al., Nature* 377 *Supp*:3-174 (1995)). Polynucleotides of the invention can also be obtained from natural sources such as genomic DNA libraries or can be synthesized using well known and commercially available techniques.

Thus, the nucleotide sequence encoding ELL2 polypeptides may be identical over its entire length to the coding sequence in SEQ ID NO:1 or may be a degenerate form of this nucleotide sequence encoding the polypeptide of SEQ ID NO:2, or may be highly identical to a nucleotide sequence that encodes the polypeptide of SEQ ID NO:2. Preferably, the polynucleotides of the invention contain a nucleotide sequence that is highly identical, at least 95% identical, with a nucleotide sequence encoding an ELL2 polypeptide or at least 95% identical with the encoding nucleotide sequence set forth in SEQ ID NO:1.

When the polynucleotides of the invention are used for the recombinant production of ELL2 polypeptide, the polynucleotide may include the coding sequence for the full-length polypeptide or a fragment thereof, by itself; the coding sequence for the full-length polypeptide or fragment in reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, or pro or prepro-protein sequence, or other fusion peptide portions. For example, a marker sequence which facilitates purification of the fused polypeptide can be encoded. In certain preferred embodiments of this aspect of the invention, the marker sequence is a hexa-histidine peptide, as provided in the pQE vector (Qiagen, Inc.) and described in Gentz *et al., Proc Natl Acad Sci USA 86*:821-824 (1989), or is an HA tag. The polynucleotide may also contain non-coding 5′ and 3′ sequences, such as transcribed, non-translated sequences, splicing and polyadenylation signals, ribosome binding sites and sequences that stabilize mRNA.

Among particularly preferred embodiments of the invention are polynucleotides encoding ELL2 polypeptides having the amino acid sequence of set out in SEQ ID NO:2 and variants thereof.

Embodiments of the invention include isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence at least 95% identical, and more preferably at least 96%, 97%, 98% or 99% identical to (a) a nucleotide sequence encoding the polypeptide having the amino acid sequence in SEQ ID NO:2; (b) a nucleotide sequence encoding the polypeptide having the amino acid sequence in SEQ ID NO:2, but lacking the N-terminal methionine; (c) a nucleotide sequence encoding the polypeptide having the amino acid sequence at positions from about 7 to about 350 in SEQ ID NO:2; (d) a nucleotide sequence encoding the polypeptide having the amino acid sequence at positions from about 50 to about 389 in SEQ ID NO:2; (e) a nucleotide sequence encoding the polypeptide having the amino acid sequence at positions from about 443 to about 474 in SEQ ID NO:2; (f) a nucleotide sequence encoding the polypeptide having the amino acid sequence at positions from about 516 to about 640 in SEQ ID NO:2; (g) a nucleotide sequence encoding the polypeptide having the amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97863; or (h) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), (c), (d), (e), (f), or (g).

As a practical matter, whether any particular nucleic acid molecule is at least 95%, 96%, 97%, 98% or 99% identical to, for instance, the nucleotide sequence shown in SEQ ID NO:1 or to the nucleotides sequence of the deposited cDNA clone can be determined conventionally using known computer programs such as the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711).

Further preferred embodiments are polynucleotides encoding ELL2, ELL2 variants that have the amino acid sequence of the ELL2 protein of SEQ ID NO:2

in which several, 5-10, 1-5, 1-3, 1-2 or 1 amino acid residues are substituted, deleted or added, in any combination.

Further preferred embodiments of the invention are polynucleotides that are at least 95% identical over their entire length to a polynucleotide encoding the ELL2 polypeptide having the amino acid sequence set out in SEQ ID NO:2, and polynucleotides which are complementary to such polynucleotides. Most highly preferred are polynucleotides that comprise a region that is at least 95% identical over their entire length to a polynucleotide encoding the ELL2 polypeptide of the human cDNA of the deposited clone and polynucleotides complementary thereto. In this regard, polynucleotides at least 96% identical over their entire length to the same are particularly preferred, and those with at least 97% are especially preferred. Furthermore, those with at least 98% are highly preferred and with at least 99% being the most preferred.

In addition, the present inventors have identified the following cDNAs related to extensive portions of SEQ ID NO:1: HPRAE28R (SEQ ID NO:9), HSBAI43R (SEQ ID NO:10), HNEAK22RA (SEQ ID NO:11), HPRTS01R (SEQ ID NO:12), HBWAL95R (SEQ ID NO:13), and HSXCR53RA (SEQ ID NO:14). In one specific embodiment, the nucleic acid molecules of the invention are not the cDNAs identified in any of SEQ ID NO:9, SEQ ID NO:10, SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, and SEQ ID NO:14.

The following public ESTs, which relate to portions of SEQ ID NO:1, have also been identified: W92650 (SEQ ID NO:15), W94585 (SEQ ID NO:16), AA243384 (SEQ ID NO:17), AA655966 (SEQ ID NO:18), N39822 (SEQ ID NO:19), AA545429 (SEQ ID NO:20), R16400 (SEQ ID NO:21), T89063 (SEQ ID NO:22), AA370048 (SEQ ID NO:23), AA375277 (SEQ ID NO:24), R12663 (SEQ ID NO:25), AA414990 (SEQ ID NO:26), AA252607 (SEQ ID NO:27), AA191245 (SEQ ID NO:28), AA524290 (SEQ ID NO:29), AA370180 (SEQ ID NO:30), Z20670 (SEQ ID NO:31), and AA009921 (SEQ ID NO:32).

The present invention is further directed to fragments of the isolated nucleic acid molecules described herein. By a fragment of an isolated nucleic acid

molecule having the nucleotide sequence of the deposited cDNA or the nucleotide sequence shown in SEQ ID NO:1 is intended fragments at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length which are useful as diagnostic probes and primers as discussed herein. Of course larger DNA fragments 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, or 2139 nt in length of the sequence shown in SEQ ID NO:1 are also useful according to the present invention as are fragments corresponding to most, if not all, of the nucleotide sequence of the cDNA clone contained in the plasmid deposited as ATCC Deposit No. 97863 or as shown in SEQ ID NO:1. By a fragment at least 20 nt in length, for example, is intended fragments which include 20 or more contiguous bases from the nucleotide sequence of the deposited cDNA or the nucleotide sequence as shown in SEQ ID NO:1.

The present invention further relates to polynucleotides that hybridize to the herein above-described sequences. In this regard, the present invention especially relates to polynucleotides which hybridize under stringent conditions to the herein above-described polynucleotides. As herein used, the term "stringent conditions" means hybridization will occur only if there is at least 95% and preferably at least 97% identity between the sequences.

Polynucleotides of the invention which are sufficiently identical to a nucleotide sequence contained in SEQ ID NO:1 or in the cDNA insert of ATCC Deposit No. 97863, may be used as hybridization probes for cDNA and genomic DNA, to isolate full-length cDNAs and genomic clones encoding ELL2 protein and to isolate cDNA and genomic clones of other genes that have a high sequence similarity to the ELL2 gene. Such hybridization techniques are known to those of skill in the art. Typically, these nucleotide sequences are 95% identical, preferably 96% identical, more preferably 97%, 98% or 99% identical to that of the referent. The probes generally will comprise at least 15 nucleotides.

Preferably, such probes will have at least 30 nucleotides and may have at least 50 nucleotides. Particularly preferred probes will range between 30 and 50 nucleotides.

In a more specific embodiment, the nucleic acid molecules of the present invention, *e.g.*, isolated nucleic acids comprising a polynucleotide having a nucleotide sequence encoding an ELL2 polypeptide or fragments thereof, are not the sequence of nucleotides, the nucleic acid molecules (*e.g.*, clones), or the nucleic acid inserts identified in one or more of the following GenBank Accession Reports: W92650, W94585, AA243384, AA655966, N39822, AA545429 R16400, T89063, AA370048, AA375277, R12663, AA414990, AA252607, AA191245, AA524290, AA370180, Z20670, and AA009921, all of which are incorporated herein by reference.

The polynucleotides and polypeptides of the present invention may be employed as research reagents and materials for discovery of treatments and diagnostics to animal and human disease.

Vectors, Host Cells, Expression

The present invention also relates to vectors which comprise a polynucleotide or polynucleotides of the present invention, and host cells which are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell-free translation systems can also be employed to produce such proteins using RNAs derived from the DNA constructs of present invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof for polynucleotides of the present invention. Introduction of polynucleotides into host cells can be effected by methods described in many standard laboratory manuals, such as Davis *et al.*, *BASIC METHODS IN MOLECULAR BIOLOGY* (1986) and Sambrook *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL*, 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989) such as calcium

phosphate transfection, DEAE-dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading, ballistic introduction or infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, *E. coli, Streptomyces* and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells.

A great variety of expression systems can be used. Such systems include, among others, chromosomal, episomal and virus-derived systems, *e.g.*, vectors derived from bacterial plasmids, from bacteriophage, from transposons, from yeast episomes, from insertion elements, from yeast chromosomal elements, from viruses such as baculoviruses, papova viruses, such as SV40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses, and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression systems may contain control regions that regulate as well as engender expression. Generally, any system or vector suitable to maintain, propagate or express polynucleotides to produce a polypeptide in a host may be used. The appropriate nucleotide sequence may be inserted into an expression system by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook *et al., MOLECULAR CLONING, A LABORATORY MANUAL (supra).*

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment appropriate secretion signals may be incorporated into the desired polypeptide. These signals may be endogenous to the polypeptide or they may be heterologous signals.

As used herein, the term "operably linked," when used in the context of a linkage between a structural gene and an expression control sequence, *e.g.*, a

promoter, refers to the position and orientation of the expression control sequence relative to the structural gene so as to permit expression of the structural gene in any host cell. For example, an operable linkage would maintain proper reading frame and would not introduce any in frame stop codons.

As used herein, the term "heterologous promoter," refers to a promoter not normally and naturally associated with the structural gene to be expressed. For example, in the context of expression of an ELL2 polypeptide, a heterologous promoter would be any promoter other than an endogenous promoter associated with the ELL2 gene in non-recombinant human chromosomes. In specific embodiments of this invention, the heterologous promoter is not a prokaryotic or bacteriophage promoter, such as the lac promoter, T3 promoter, or T7 promoter. In other embodiments, the heterologous promoter is a eukaryotic promoter.

In other embodiments this invention provides an isolated nucleic acid molecule comprising an ELL2 structural gene operably linked to a heterologous promoter. As used herein, the term "an ELL2 structural gene" refers to a nucleotide sequence at least 95% identical to one of the following nucleotide sequences:

(a) a nucleotide sequence encoding the ELL2 polypeptide having the complete amino acid sequence in SEQ ID NO:2;

(b) a nucleotide sequence encoding the ELL2 polypeptide having the amino acid sequence at positions 2-640 in SEQ ID NO:2;

(c) a nucleotide sequence encoding the ELL2 polypeptide having the complete amino acid sequence encoded by the cDNA clone contained in ATCC Deposit No. 97863; or

(d) a nucleotide sequence complementary to any of the nucleotide sequences in (a), (b), or (c).

In more preferred embodiments, the ELL2 structural gene is 96%, 97%, 98%, 99%, or 100% identical to one or more of nucleotide sequences (a), (b), or (c) *supra*.

This invention also provides an isolated nucleic acid molecule comprising an ELL2 structural gene operably linked to a heterologous promoter, wherein said isolated nucleic acid molecule does not encode a fusion protein comprising the ELL2 structural gene or a fragment thereof. In particular embodiments the isolated nucleic acid molecule does not encode a beta-galactosidase-ELL2 fusion protein.

This invention further provides an isolated nucleic acid molecule comprising an ELL2 structural gene operably linked to a heterologous promoter, wherein said isolated nucleic acid molecule is capable of expressing an ELL2 polypeptide when used to transform an appropriate host cell. In particular embodiments, the ELL2 polypeptide does not contain and is not covalently linked to an amino acid sequence encoded by the 5' untranslated portion of the ELL2 gene, *e.g.*, nucleotides 1-93 of SEQ ID NO:1, or a fragment thereof.

This invention also provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence encoding a ELL2 polypeptide having the amino acid sequence of SEQ ID NO:2, wherein said isolated nucleic acid molecule does not contain a nucleotide sequence at least 90% identical to the 3' untranslated region of SEQ ID NO:1 (nucleotides 2014-2139), or a fragment of the 3' untranslated region greater than 25, 50, 75, 100, or 125 bp in length. In other embodiments, said isolated nucleic acid molecule does not contain a nucleotide sequence at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the 3' untranslated region of SEQ ID NO:1 (nucleotides 2014-2139).

This invention further provides an isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence encoding a ELL2 polypeptide having the amino acid sequence of SEQ ID NO:2, wherein said isolated nucleic acid molecule does not contain a nucleotide sequence at least 90% identical to the 5' untranslated region of SEQ ID NO:1 (nucleotides 1-93), or a fragment of the 5' untranslated region greater than 25, 35, 45, 55, 65, 75, 85, or 90 bp. In other embodiments, said isolated nucleic

acid molecule does not contain a nucleotide sequence at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the 5' untranslated region of SEQ ID NO:1 (nucleotides 1-93).

In addition, the present invention further includes novel expression vectors comprising operator and promoter elements operatively linked to nucleotide sequences encoding a protein of interest. One example of such a vector is pHE4-5 which is described in detail below.

As summarized in FIG. 4 and FIG. 5, components of the pHE4-5 vector (SEQ ID NO:33) include: 1) a neomycinphosphotransferase gene as a selection marker, 2) an *E. coli* origin of replication, 3) a T5 phage promoter sequence, 4) two *lac* operator sequences, 5) a Shine-Delgarno sequence, 6) the lactose operon repressor gene (*lac*Iq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences were made synthetically. Synthetic production of nucleic acid sequences is well known in the art. CLONTECH 95/96 Catalog, pages 215-216, CLONTECH, 1020 East Meadow Circle, Palo Alto, CA 94303. A nucleotide sequence encoding ELL2 (SEQ ID NO:1), is operatively linked to the promoter and operator by inserting the nucleotide sequence between the *Nde*I and *Asp*718 sites of the pHE4-5 vector.

As noted above, the pHE4-5 vector contains a *lac*Iq gene. *Lac*Iq is an allele of the *lac*I gene which confers tight regulation of the *lac* operator. Amann, E. *et al.*, *Gene 69*:301-315 (1988); Stark, M., *Gene 51*:255-267 (1987). The *lac*Iq gene encodes a repressor protein which binds to *lac* operator sequences and blocks transcription of down-stream (*i.e.*, 3') sequences. However, the *lac*Iq gene product dissociates from the *lac* operator in the presence of either lactose or certain lactose analogs, *e.g.*, isopropyl B-D-thiogalactopyranoside (IPTG). ELL2 thus is not produced in appreciable quantities in uninduced host cells containing the pHE4-5 vector. Induction of these host cells by the addition of an agent such as IPTG, however, results in the expression of the ELL2 coding sequence.

The promoter/operator sequences of the pHE4-5 vector (SEQ ID NO:33) comprise a T5 phage promoter and two *lac* operator sequences. One operator is located 5' to the transcriptional start site and the other is located 3' to the same site. These operators, when present in combination with the *lac*Iq gene product, confer tight repression of down-stream sequences in the absence of a *lac* operon inducer, *e.g.*, IPTG. Expression of operatively linked sequences located down-stream from the *lac* operators may be induced by the addition of a *lac* operon inducer, such as IPTG. Binding of a *lac* inducer to the *lac*Iq proteins results in their release from the *lac* operator sequences and the initiation of transcription of operatively linked sequences. *Lac* operon regulation of gene expression is reviewed in Devlin, T., TEXTBOOK OF BIOCHEMISTRY WITH CLINICAL CORRELATIONS, 4th Edition (1997), pages 802-807.

The pHE4 series of vectors contain all of the components of the pHE4-5 vector except for the ELL2 coding sequence. Features of the pHE4 vectors include optimized synthetic T5 phage promoter, *lac* operator, and Shine-Delgarno sequences. Further, these sequences are also optimally spaced so that expression of an inserted gene may be tightly regulated and high level of expression occurs upon induction.

Among known bacterial promoters suitable for use in the production of proteins of the present invention include the *E. coli lac*I and *lac*Z promoters, the T3 and T7 promoters, the *gpt* promoter, the lambda PR and PL promoters and the *trp* promoter. Suitable eukaryotic promoters include the CMV immediate early promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous Sarcoma Virus (RSV), and metallothionein promoters, such as the mouse metallothionein-I promoter.

The pHE4-5 vector also contains a Shine-Delgarno sequence 5' to the AUG initiation codon. Shine-Delgarno sequences are short sequences generally located about 10 nucleotides up-stream (*i.e.*, 5') from the AUG initiation codon.

These sequences essentially direct prokaryotic ribosomes to the AUG initiation codon.

Thus, the present invention is also directed to expression vector useful for the production of the proteins of the present invention. This aspect of the invention is exemplified by the pHE4-5 vector (SEQ ID NO:33).

If the ELL2 polypeptide is to be expressed for use in screening assays, generally, it is preferred that the polypeptide be produced at the surface of the cell. In this event, the cells may be harvested prior to use in the screening assay. If ELL2 polypeptide is secreted into the medium, the medium can be recovered in order to recover and purify the polypeptide; if produced intracellularly, the cells must first be lysed before the polypeptide is recovered.

ELL2 polypeptides can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography is employed for purification. Well known techniques for refolding proteins may be employed to regenerate active conformation when the polypeptide is denatured during isolation and or purification.

Diagnostic Assays

This invention also relates to the use of ELL2 polynucleotides for use as diagnostic reagents. Detection of a mutated form of ELL2 gene associated with a dysfunction will provide a diagnostic tool that can add to or define a diagnosis of a disease or susceptibility to a disease which results from under-expression, over-expression or altered expression of ELL2. Individuals carrying mutations in the ELL2 gene may be detected at the DNA level by a variety of techniques.

Nucleic acids for diagnosis may be obtained from a subject's cells, such as from blood, urine, saliva, tissue biopsy or autopsy material. The genomic DNA

may be used directly for detection or may be amplified enzymatically by using PCR or other amplification techniques prior to analysis. RNA or cDNA may also be used in similar fashion. Deletions and insertions can be detected by a change in size of the amplified product in comparison to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to labeled ELL2 nucleotide sequences. Perfectly matched sequences can be distinguished from mismatched duplexes by RNase digestion or by differences in melting temperatures. DNA sequence differences may also be detected by alterations in electrophoretic mobility of DNA fragments in gels, with or without denaturing agents, or by direct DNA sequencing. See, *e.g.*, Myers *et al.*, *Science 230*:1242 (1985). Sequence changes at specific locations may also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method. See Cotton *et al.*, *Proc. Natl. Acad. Sci. USA 85*:4397-4401 (1985).

The diagnostic assays offer a process for diagnosing or determining a susceptibility to neoplastic disorders through detection of mutation in the ELL2 gene by the methods described.

In addition, neoplastic disorders, can be diagnosed by methods comprising determining from a sample derived from a subject an abnormally decreased or increased level of ELL2 polypeptide or ELL2 mRNA. Decreased or increased expression can be measured at the RNA level using any of the methods well known in the art for the quantitation of polynucleotides, such as, for example, PCR, RT-PCR, RNase protection, Northern blotting and other hybridization methods. Assay techniques that can be used to determine levels of a protein, such as an ELL2 protein, in a sample derived from a host are well-known to those of skill in the art. Such assay methods include radioimmunoassays, competitive-binding assays, Western Blot analysis and ELISA assays.

Additionally, methods are provided for diagnosing or determining a susceptibility of an individual to neoplastic disorders, comprising (a) assaying ELL2 protein gene expression level in mammalian cells or body fluid; and (b) comparing said ELL2 protein gene expression level with a standard ELL2 protein

gene expression level whereby an increase or decrease in said ELL2 gene expression level over said standard is indicative of an increased or decreased susceptibility to a neoplastic disorder.

Chromosome Assays

The nucleotide sequences of the present invention are also valuable for chromosome identification. The sequence is specifically targeted to and can hybridize with a particular location on an individual human chromosome. For example, ELL2 sequences have been found by fluorescent *in situ* hybridization to bind to human chromosomes at 1 q21 and 5 q15. The inventors have further found that the hybrization signal from the 1 q21 locus is significantly stronger than that at 5 q15.

The mapping of relevant sequences to chromosomes according to the present invention is an important first step in correlating those sequences with genes associated disease. Once a sequence has been mapped to a precise chromosome location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on-line through Johns Hopkins, University Welch Medical Library). The relationship between genes and diseases that have been mapped to the same chromosomal region are then identified through linkage analysis (coinheritence of physically adjacent genes).

The differences in the cDNA or genomic sequence between affected and unaffected individuals can also be determined. If a mutation is observed in some or all of the affected individuals but not in any normal individuals, then the mutation is likely to be the causative agent of the disease.

Antibodies

The polypeptides of the invention or their fragments or analogs thereof, or cells expressing them can also be used as immunogens to produce antibodies immunospecific for the ELL2 Polypeptides. By "immunospecific" is meant that the antibodies have affinities for the polypeptides of the invention that are

substantially greater in their affinities for related polypeptides such as the analogous proteins of the prior art.

Antibodies generated against the ELL2 polypeptides can be obtained by administering the polypeptides or epitope-bearing fragments, analogs or cells to an animal, preferably a nonhuman, using routine protocols. For preparation of monoclonal antibodies, any technique which provides antibodies produced by continuous cell line cultures can be used. Examples include the hybridoma technique (Kohler, G. and Milstein, C., *Nature 256*:495-497 (1975)), the trioma technique, the human B-cell hybridoma technique (Kozbor *et al., Immunology Today 4*:72 (1983)) and the EBV-hybridoma technique (Cole *et al.,* MONOCLONAL ANTIBODIES AND CANCER THERAPY, pp. 77-96, Alan R. Liss, Inc., 1985).

Techniques for the production of single chain antibodies (U.S. Patent No. 4,946,778) can also be adapted to produce single chain antibodies to polypeptides of this invention. Also, transgenic mice, or other organisms including other mammals, may be used to express humanized antibodies.

The above-described antibodies may be employed to isolate or to identify clones expressing the polypeptide or to purify the polypeptides by affinity chromatography.

Antibodies against ELL2 polypeptides may also be employed to treat neoplastic disorders, among others.

Screening Assays

The ELL2 of the present invention may be employed in a screening process for compounds which bind the protein and which activate (agonists) or inhibit activation of (antagonists) the polypeptide of the present invention. Thus, polypeptides of the invention may also be used to assess the binding of small molecule substrates and ligands in, for example, cells, cell-free preparations, chemical libraries, and natural product mixtures. These substrates and ligands may

be natural substrates and ligands or may be structural or functional mimetics. See Coligan *et al., Current Protocols in Immunology 1(2)*:Chapter 5 (1991).

By "agonist" is intended naturally occurring and synthetic compounds capable of enhancing an ELL2 activity (*e.g.*, promoting ELL2 catalyzed increased rate of RNA elongation during transcription). By "antagonist" is intended naturally occurring and synthetic compounds capable of inhibiting an ELL2 activity.

ELL2 proteins are associated with many pathologies. Accordingly, it is desirous to find compounds and drugs which stimulate ELL2 on the one hand and which can inhibit the function of ELL2 on the other hand. In general, agonists are employed for therapeutic and prophylactic purposes including the treatment of ceratin types of neoplastic disorders. For example, the ELL gene was originally isolated as a gene that undergoes frequent translocations in acute myeloid leukemia (Thirman, M.J. *et al., Proc. Natl. Acad. Sci. U.S.A. 91*:12110-12114 (1994), Mitani, K. *et al., Blood 85*:2017-2024 (1995)). ELL2 gene translocations could result in similar disorders which may be treated by enhancing ELL2 activity, *e.g.*, by administration of an ELL2 agonist or an ELL2 polypeptide. Further, overexpression of ELL leads to the transformation of fibroblasts. Thus, antagonists of ELL2 activity may be employed for a variety of therapeutic and prophylactic purposes including the treatment of certain types of neoplastic disorders associated with overexpression of ELL2.

In general, such screening procedures involve producing appropriate cells which express the polypeptide of the present invention on the surface thereof. Such cells include cells from mammals, yeast, *Drosophila* or *E. coli.* Cells expressing the protein (or cell membrane containing the expressed protein) are then contacted with a test compound to observe binding, or stimulation or inhibition of a functional response.

The assays may simply test binding of a candidate compound wherein adherence to the cells bearing the protein is detected by means of a label directly or indirectly associated with the candidate compound or in an assay involving

competition with a labeled competitor. Further, these assays may test whether the candidate compound results in a signal generated by activation of the protein, using detection systems appropriate to the cells bearing the protein at their surfaces. Inhibitors of activation are generally assayed in the presence of a known agonist and the effect on activation by the agonist by the presence of the candidate compound is observed. Standard methods for conducting such screening assays are well understood in the art.

Examples of potential ELL2 protein antagonists include antibodies or, in some cases, oligonucleotides or proteins which are closely related to the ligands of the ELL2 protein, *e.g.*, a fragment of the ligand, or small molecules which bind to the protein but do not elicit a response, so that the activity of the protein is prevented. Such ligands include other molecules involved in the process of transcription and SH3 domains.

Prophylactic and Therapeutic Methods

This invention provides methods of treating an abnormal conditions related to both an excess of and insufficient amounts of ELL2 protein activity.

If the activity of ELL2 protein is in excess, several approaches are available. One approach comprises administering to a subject an inhibitor compound (antagonist) as hereinabove described along with a pharmaceutically acceptable carrier in an amount effective to inhibit activation by blocking binding of ligands to the ELL2 protein, or by inhibiting a second signal, and thereby alleviating the abnormal condition.

In another approach, soluble forms of ELL2 polypeptides still capable of binding the ligand in competition with endogenous ELL2 protein may be administered. Typical embodiments of such competitors comprise fragments of the ELL2 polypeptide.

In still another approach, expression of the gene encoding endogenous ELL2 protein can be inhibited using expression blocking techniques. Known such techniques involve the use of antisense sequences, either internally generated or

separately administered. See, for example, O'Connor, *J. Neurochem* 56:560 (1991) in OLIGODEOXYNUCLEOTIDES AS ANTISENSE INHIBITORS OF GENE EXPRESSION, CRC Press, Boca Raton, FL (1988). Alternatively, oligonucleotides which form triple helices with the gene can be supplied. See, for example, Lee *et al., Nucleic Acids Res 6:*3073 (1979); Cooney *et al., Science 241*:456 (1988); Dervan *et al., Science 251*:1360 (1991). These oligomers can be administered *per se* or the relevant oligomers can be expressed *in vivo.*

For treating abnormal conditions related to an under-expression of ELL2 and its activity, several approaches are also available. One approach comprises administering to a subject a therapeutically effective amount of a compound which activates ELL2, *i.e.*, an agonist as described above, in combination with a pharmaceutically acceptable carrier, to thereby alleviate the abnormal condition. Alternatively, gene therapy may be employed to effect the endogenous production of ELL2 by the relevant cells in the subject. For example, a polynucleotide of the invention may be engineered for expression in a replication defective retroviral vector, as discussed above. The retroviral expression construct may then be isolated and introduced into a packaging cell transduced with a retroviral plasmid vector containing RNA encoding a polypeptide of the present invention such that the packaging cell now produces infectious viral particles containing the gene of interest. These producer cells may be administered to a subject for engineering cells *in vivo* and expression of the polypeptide *in vivo*. For overview of gene therapy, see Strachan, T. & Read A.P., Chapter 20, "Gene Therapy and Other Molecular Genetic-based Therapeutic Approaches," (and references cited therein) in HUMAN MOLECULAR GENETICS, BIOS Scientific Publishers Ltd. (1996).

Formulation and Administration

Peptides, such as the soluble form of ELL2 polypeptides, and agonists and antagonist peptides or small molecules, may be formulated in combination with a suitable pharmaceutical carrier. Such formulations comprise a therapeutically effective amount of the polypeptide or compound, and a pharmaceutically

acceptable carrier or excipient. Such carriers include but are not limited to, saline, buffered saline, dextrose, water, glycerol, ethanol, and combinations thereof. Formulation should suit the mode of administration, and is well within the skill of the art. The invention further relates to pharmaceutical packs and kits comprising one or more containers filled with one or more of the ingredients of the aforementioned compositions of the invention.

Polypeptides and other compounds of the present invention may be employed alone or in conjunction with other compounds, such as therapeutic compounds.

Preferred forms of systemic administration of the pharmaceutical compositions include injection, typically by intravenous injection. Other injection routes, such as subcutaneous, intramuscular, or intraperitoneal, can be used. Alternative means for systemic administration include transmucosal and transdermal administration using penetrants such as bile salts or fusidic acids or other detergents. In addition, if properly formulated in enteric or encapsulated formulations, oral administration may also be possible. Administration of these compounds may also be topical and/or localized, in the form of salves, pastes, gels and the like.

The dosage range required depends on the choice of peptide, the route of administration, the nature of the formulation, the nature of the subject's condition, and the judgment of the attending practitioner. Suitable dosages, however, are in the range of 0.1-100 μg/kg of subject. Wide variations in the needed dosage, however, are to be expected in view of the variety of compounds available and the differing efficiencies of various routes of administration. For example, oral administration would be expected to require higher dosages than administration by intravenous injection. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art.

Polypeptides used in treatment can also be generated endogenously in the subject, in treatment modalities often referred to as "gene therapy" as described above. Thus, for example, cells from a subject may be engineered with a

polynucleotide, such as a DNA or RNA, to encode a polypeptide *ex vivo*, and for example, by the use of a retroviral plasmid vector. The cells are then introduced into the subject.

Example

The example below is carried out using standard techniques, which are well known and routine to those of skill in the art, except where otherwise described in detail. The example illustrates, but does not limit the invention.

Example 1.

Material and Methods

Cloning and Expression of Wild Type and Mutant ELL2

A DNA fragment including *ELL2* coding sequences was obtained by PCR amplification of a Lambda Zap human fetal heart library (Stratagene) using a 5' primer (5'-CAATTAACCCTCATAAAGGGAAC-3') (SEQ ID NO:3) identical to a sequence in the Lambda Zap vector and a 3' antisense primer (5'-CAAAGTTTCACCTTTTAGAATCTAGAGCAACTC-3') (SEQ ID NO:4) corresponding to a sequence in the 3'-untranslated region of the *ELL2* gene. The construct for expression of histidine-tagged ELL2 in bacteria was prepared in two steps. First, a DNA fragment encoding ELL2 amino acids 11-640 was generated by PCR amplification of the original *ELL2* ORF-containing PCR product using the *ELL2*- specific primers:

5'-GAGGTGTCGACGAGGAGCAGCGCTATGGGCTGTCGTGCGGAC-3' (SEQ ID NO:5) and

5'-GTGTGGATCCTCATCACTAGGACCATGACTCTGCTTGCTGTTG-3' (SEQ ID NO:6) and was introduced into the *Sal*I and *Bam*HI sites of M13mpET (Tan, S. *et al., BioTechniques 16*:824-828 (1994)). An expression vector

containing the entire *ELL2* ORF was then generated by oligonucleotide-directed mutagenesis (Kunkel, T.A., *Proc. Natl. Acad. Sci. U.S.A. 82*:488-492 (1985)) with the Muta-Gene M13 *in vitro* mutagenesis kit (Bio-Rad) and confirmed by DNA sequencing; N- and C-terminal ELL2 deletion mutants were constructed by the same procedure. Wild type and mutant ELL2 proteins were expressed in *E. coli,* purified from guanidine-solubilized inclusion bodies by nickel affinity chromatography, and renatured as described by Shilatifard, A. *et al., Science 271*:1873-1876 (1996). Where indicated, the ELL2 protein was further purified by preparative SDS-polyacrylamide gel electrophoresis (Shilatifard, A. *et al., Science 271*:1873-1876 (1996)). The human ELL protein was expressed in *E. coli* and purified as described (Shilatifard, A. *et al., Science 271*:1873-1876 (1996)).

Tissue distribution of ELL2 and ELL mRNAs

A human multiple tissue northern blot (MTN1, Clontech) was probed with PCR generated, *ELL2*- and *ELL*-specific probes chosen from a region of sequence that was most divergent between the two genes. The *ELL*-specific probe contained sequences encoding amino acids 317-621, and the *ELL2*-specific probe contained sequences encoding amino acids 327-474. Probes were labeled with [α-^{32}P]dCTP by random priming performed according to the manufacturer's instructions (Rediprime kit, Amersham). The blot was prehybridized in 10 ml of Hybrisol I solution (Oncor) for 3 hr at 42°C. Probe DNA was denatured and added to hybridization solution at 10^6 cpm/ml of solution. Hybridization was carried out at 42°C overnight. The blot was washed 10 min in 2xSSC/0.1% SDS at room temperature, 15 min in 0.2xSSC/0.1% SDS at 45°C, 10 min in 0.1xSSC/0.1% SDS at 55°C, and then exposed to film (Hyperfilm-MP, Amersham) overnight at -80°C.

Assay of the effects of ELL2 on elongation by RNA polymerase II during synthesis of promoter-independent and promoter-dependent transcripts

ELL2 and ELL have similar effects on elongation by RNA polymerase II during synthesis of promoter-independent and promoter-dependent transcripts. 10% SDS-PAGE of recombinant ELL2 (rELL2) and ELL (rELL), purified by nickel chromatography and preparative SDS-PAGE. Proteins were visualized by silver staining. Effects of ELL2 and ELL on the kinetics of promoter-dependent transcription. Preinitiation complexes were assembled at the AdML promoter with recombinant TBP, TFIIB, TFIIE, TFIIF, and purified rat TFIIH and RNA polymerase II as described (Shilatifard, A. *et al., Science 271*:1873-1876 (1996)). Transcription was initiated by addition of 50 μM ATP, 50 μM GTP, 2 μM UTP, 10 μCi of [α-^{32}P]CTP (>400 Ci/mmol, Amersham) and 7 mM $MgCl_2$. After 10 min at 28°C, 100 μM nonradioactive CTP was added to reaction mixture and short transcripts were chased in the absence or presence of ~50 ng SDS-PAGE purified rELL2 or rELL for the times indicated. Transcripts were analyzed by electrophoresis through a 6% polyacrylamide, 7.0 M urea gel.

The following procedure was used to determine the effects that ELL2 and ELL have on the kinetics of promoter-independent transcription. SDS/PAGE purified histidine-tagged ELL2 and ELL proteins were renatured and assayed in pulse-chase reactions using the oligo(dC)-tailed template pCpGR220 S/P/X. Reactions contained ~0.01 units of RNA polymerase II, 100 ng of pCpGR220S/P/X, and ~50 ng rELL2 or ~50 ng ELL and were performed essentially as described (Shilatifard, A. *et al., Science 271*:1873-1876 (1996)). The control reaction (mock) contained an identically prepared fraction from uninfected JM109(DE3) cells.

Localization of the ELL2 elongation activation domain

Wild type ELL2 and ELL2 mutants were expressed in *E. coli* and purified by nickel affinity chromatography as described above. Approximately 50 ng of each protein (in a maximum volume of 50 μl) was renatured and assayed as described (Shilatifard, A. *et al., Science 271*:1873-1876 (1996)) for its ability to

stimulate synthesis of the 135 nucleotide transcript from the T-less cassette of oligo(dC)-tailed template pCpGR220 S/P/X. Reactions containing ~0.01 unit RNA polymerase II, 100 ng template, and the indicated ELL2 proteins were incubated at 28 °C for 5 min in the presence of 50 μM ATP, 50 μM GTP, 1.8 μM CTP, and 10 μCi [α-^{32}P]CTP. The control reaction (mock) contained an identically prepared fraction from uninfected JM109(DE3) cells.

Results

Identification of Human ELL2

Searches of the Human Genome Sciences and GenBank EST databases identified multiple overlapping ESTs that formed a contig spanning a predicted *ELL2* ORF similar in sequence to the ORF of the human *ELL* gene (FIG. 2). An ~1.9 kb DNA fragment containing the entire predicted *ELL2* ORF was obtained by PCR amplification of a human fetal heart library and sequenced. The *ELL2* ORF encodes a 640 amino acid protein with a calculated molecular mass of 72,354 Da. As determined by the BESTFIT program of the Genetics Computer Group (GCG, Madison, WI) package (Genetics Computer Group, *Program Manual for the GCG Package, Version 8* (Madison, WI) (1994)), ELL2 is 49% identical and 66% similar to ELL (alignment score ~64 SD).

Expression of ELL2 and ELL in Human Cells

To investigate the expression of ELL2 and ELL in human cells, Northern blots containing poly A$^+$ RNA from various human tissues were hybridized with ELL2- and ELL-specific probes. Consistent with previous studies (Thirman, M.J. *et al., Proc. Natl. Acad. Sci. U.S.A. 91*:12110-12114 (1994), Mitani, K. *et al., Blood 85*:2017-2024 (1995)), the ELL-specific probe hybridized to two mRNA species of ~4.4 kb and ~2.7 kb. The ELL2-specific probe hybridized to two mRNA species of ~7 kb and ~ 4.1 kb. At present, it is not clear whether the ~7 kb and 4.1 kb ELL2 mRNAs are alternatively processed forms or the products of

closely related genes. The results of Northern blot analysis indicate that both ELL2 and ELL mRNAs are expressed in many of the same tissues, *i.e.*, both ELL2 and ELL mRNAs are expressed at the highest levels in pancreas, skeletal muscle, placenta, and heart, and at lower levels in lung, and brain. Unlike ELL, ELL2 is expressed at high levels in liver, but at nearly undetectable levels in kidney. Notably, the ratio of ELL2 and ELL mRNAs, as well as the ratios of the two different forms of each mRNA, exhibit tissue to tissue variation.

ELL2 and ELL Possess Similar Transcriptional Activities

ELL is capable of potently stimulating the overall rate of RNA chain elongation by RNA polymerase II (Shilatifard, A. *et al., Science 271*:1873-1876 (1996)). To demonstrate that ELL2 is also capable of stimulating elongation by RNA polymerase II, a DNA fragment containing the *ELL2* ORF was introduced into a bacteriophage M13 expression vector under control of the T7 RNA polymerase promoter (Tan, S. *et al., BioTechniques 16*:824-828 (1994)) and expressed in *E. coli* with an N-terminal histidine tag. The recombinant ELL2 protein was purified to homogeneity from guanidine-solubilized inclusion bodies by nickel affinity chromatography and preparative SDS-gel electrophoresis, and tested for its ability to stimulate elongation.

ELL2 is an RNA polymerase II elongation factor with functional properties similar to those of ELL. The abilities of ELL2 and ELL to stimulate elongation were compared during either promoter-specific transcription carried out in the presence of the general initiation factors or promoter-independent transcription carried out using an oligo(dC)-tailed template assay, in the absence of auxiliary transcription factors.

To compare the abilities of ELL2 and ELL to stimulate the rate of elongation of promoter-specific transcripts, preinitiation complexes were assembled by preincubation of purified RNA polymerase II, TBP, TFIIB, TFIIE, TFIIF, and TFIIH with a DNA template containing the AdML promoter. Short, highly radioactive transcripts were then synthesized during a brief pulse carried

out in the presence of ATP, GTP, UTP, and a limiting concentration of [α-^{32}P]CTP. These short, promoter-specific transcripts were then chased into full-length runoff transcripts in the presence of an excess of nonradioactive CTP and in the presence or absence of approximately equivalent levels of recombinant ELL2 or ELL. Comparison of the kinetics of accumulation of full-length runoff transcripts reveals that ELL2 and ELL have similar effects on the rate of elongation of promoter-specific transcripts by RNA polymerase II.

An oligo(dC)-tailed template assay was used to compare the abilities of ELL2 and ELL to stimulate the rate of elongation of promoter-independent transcripts. Briefly, transcription was initiated by addition of RNA polymerase II to reaction mixtures containing the oligo(dC)-tailed template pCpGR220 S/P/X (Rice, G.A. *et al., Proc. Natl. Acad. Sci. U.S.A. 88*:4245-4249 (1991)), ATP, GTP, and [α-^{32}P]CTP. Under these conditions, RNA polymerase II synthesizes ~135 nucleotide transcripts on the T-less cassette of pCpGR220 S/P/X. These highly radioactive transcripts were then chased into longer RNAs with UTP and an excess of nonradioactive CTP, in the presence or absence of approximately equivalent levels of recombinant ELL2 or ELL. Transcripts synthesized in the presence of either ELL2 or ELL were substantially longer than transcripts synthesized in their absence; we note that many transcripts synthesized in the presence of ELL2 and ELL appear to be plasmid length. In addition, comparison of the kinetics of accumulation of long transcripts and of the distribution of RNA intermediates reveals that ELL2 and ELL have similar effects on elongation of transcripts synthesized by RNA polymerase II in the absence of auxiliary transcription factors on the oligo(dC)-tailed pCpGR220 S/P/X template.

Localization of the ELL2 Elongation Activation Domain

Comparison of the ELL2 and ELL ORFs revealed three conserved regions (FIG. 2 and FIG. 3): an N-terminal region (region 1) between ELL2 residues 7 and 350, a short lysine-rich region (region 2) between ELL2 residues 443 and 474, and a C-terminal region (region 3) between ELL2 residues 516 and 640.

Although neither ELL2 nor ELL have obvious structural features such as zinc finger, leucine zipper, or helix-turn-helix motifs commonly found in transcription factors, a TBLASTN search of the GenBank database revealed that conserved region 3 of ELL2 and ELL exhibits striking similarity to the ZO-1 binding domain of occludin (SEQ ID NO:8) (Furuse, M. *et al., J. Cell. Bio. 127*:1617-1626 (1994)), an integral membrane protein found at tight junctions (Furuse, M. *et al., J. Cell. Bio. 123*:1777-1788 (1993)). As determined by the BESTFIT program of the Genetics Computer Group (GCG, Madison, WI) package (Genetics Computer Group, *Program Manual for the GCG Package, Version 8* (Madison, WI) (1994)), the C-terminus of ELL2 and the ZO-1 binding domain of occludin are 33% identical and 61% similar (alignment score ~17 SD) over a 112 amino acid region. In addition, ELL2 and ELL each contain a proline-rich, non-conserved region that bridges conserved regions 1 and 2. The ELL2 proline-rich region includes several PXXP motifs that are potential binding sites for SH3 domains (Yu, H. *et al., Cell 76*:933-945 (1996)).

To assess the functional significance of the regions conserved between ELL2 and ELL and to localize the ELL2 elongation activation domain, a series of ELL2 deletion mutants was constructed (FIG. 3), expressed in *E. coli,* purified, and tested for transcriptional activity using the oligo(dC)-tailed template assay. The results of these experiments localize the ELL2 elongation activation domain to sequences in conserved region 1 between residues 50 and 389. ELL2 deletion mutants Δ194-640, Δ100-640, and Δ50-194, which each lack significant portions of region 1, had significantly reduced transcriptional activities. In contrast, ELL2 deletion mutants Δl-10, Δ1-50, Δ499-640, and Δ389-640 all exhibited near wild type levels of activity.

Discussion

Here we report identification and characterization of ELL2, a novel RNA polymerase II elongation factor similar to previously characterized elongation factor ELL (Shilatifard, A. *et al., Science 271*:1873-1876 (1996)). ELL2 is the

newest addition to a growing list of biochemically defined cellular proteins that are capable of regulating the activity of the RNA polymerase II elongation complex. This list now includes six cellular elongation factors: SII, P-TEFb, TFIIF, Elongin (SIII), ELL, and ELL2, which fall into two distinct functional classes (Reines, D. *et al., Trends. Biochem. Sci. 21*:351-355 (1996)).

SII and P-TEFb were shown previously to prevent RNA polymerase II from arresting transcription prematurely. SII protects RNA polymerase II from arrest at a variety of transcriptional impediments, including specific DNA sequences that act as intrinsic arrest sites and some DNA bound proteins and drugs. SII promotes passage of RNA polymerase II through these transcriptional impediments by a mechanisms involving reiterative endonucleolytic cleavage and re-extension of nascent transcripts held in the polymerase site (Reines, D., in *Transcription: Mechanisms and Regulations*, eds. Conaway, R.C. & Conaway, J.W. (Raven Press, New York) (1994), pp. 263-278). P-TEFb promotes passage of RNA polymerase II through DRB-sensitive arrest sites within a few hundred nucleotides of promoters, by a mechanism that may involve phosphorylation of the RNA polymerase II CTD (Marshall, N.F. & Price, D.H., *J. Biol. Chem. 270*:12335-12338 (1995), Marshall, N.F. *et al., J. Biol. Chem. 271*:27176-27183 (1996)). TFIIF, Elongin (SIII), and ELL were all shown previously to increase the overall rate of elongation by RNA polymerase II by decreasing the frequency or duration of transient pausing by the enzyme at many sites along DNA templates (Shilatifard, A. *et al., Science 271*:1873-1876 (1996), Price, D.H. *et al., Mol. Cell. Biol. 9*:1465-1475 (1989), Aso, T. *et al., Science 269*:1439-1443 (1995)). Neither TFIIF, Elongin (SIII), nor ELL is capable of releasing RNA polymerase II from SII- or DRB-sensitive arrest sites.

As we have shown here, ELL2 regulates the activity of the RNA polymerase II elongation complex by a mechanism more closely resembling those of TFIIF, Elongin (SIII), and ELL. ELL2 appears to increase the overall rate of elongation by RNA polymerase II during both promoter-dependent and

-independent transcription. In contrast to SII, ELL2 does not release RNA polymerase II from arrest or promote the nascent transcription cleavage reaction.

Although ELL2 and ELL are related proteins, they do not share sequence similarity throughout their entire ORFs. Alignment of their ORFs revealed that ELL2 and ELL share three regions of high homology: an N-terminal region between ELL2 residues 7 and 353, a short lysine-rich region between ELL2 residues 443-474, and a C-terminal region between ELL2 residues 516-640. Structure-function analysis reveals that ELL2 transcriptional activity resides in conserved region 1 in the ELL2 N-terminus. Neither the conserved lysine-rich region 2 nor the conserved C-terminal region 3 is required for ELL2 transcriptional activity. The functions of regions 2 and 3 are presently unknown.

A homology search of the GenBank database revealed that conserved region 3 of ELL2 and ELL bears a striking resemblance to the ZO-1 binding domain of occludin (Furuse, M. *et al., J. Cell. Bio. 127*:1617-1626 (1994)), an integral membrane protein localized at tight junctions in mammalian cells (Furuse, M. *et al., J. Cell. Bio. 123*:1777-1788 (1993)). ZO-1 is a member of the family of membrane-associated guanylate kinase homologs (MAGUKs) believed to be important in signal transduction originating from sites of cell-cell contact (Willott, E. *et al., Proc. Natl. Acad. Sci. U.S.A. 90*:7834-7838 (1993)). The founding member of the MAGUK family of putative signaling proteins is the product of the *lethal(1)discs large-1 (dlg)* tumor suppressor gene of *Drosophila* (Woods, D.F. & Bryant, P.J., *Cell 66*:541-464 (1994)). Other members of the MAGUK family include ZO-2, a second tight junction protein (Jesaitis, L.A. & Goodenough, D.A., *J. Cell. Biol. 124*:949-961 (1994)), PSD-95/SAP-90, which localizes to synaptic junctions (Kistner, U. *et al., J. Biol. Chem. 268*:4580-4583 (1993)), p55, which participates in erythrocyte membrane-cytoskeletal interactions (Ruff, P. *et al., Proc. Natl. Acad. Sci. U.S.A. 88*:6595-6599 (1991)), and *hdlg*, a human homolog of *Drosophila* dlg (Lue, R.A. *et al., Proc. Natl. Acad. Sci. U.S.A. 91*:9818-9822 (1994)). Recently, ZO-1. which is found exclusively in the cytosol of contact-inhibited cultured cells, was found to translocate to the nucleus in subconfluent

cells, suggesting that ZO-1 is involved in signaling pathways controlled by cell-cell contact (Kistner, U. *et al., J. Biol. Chem. 268*:4580-4583 (1993)). Intriguingly, the intracellular localization of the product of the von Hippel-Lindau tumor suppressor gene, which has been shown to interact with and negatively regulate the B and C regulatory subunits of Elongin, is similarly regulated by cell density (Lee, S. *et al., Proc. Natl. Acad. Sci. U.S.A. 93*:1770-1775 (1996)). Whether conserved region 3 of ELL2 or ELL is capable of interacting with ZO-1 is presently unknown. However, ELL2 and ELL could be regulated *via* a signal transduction pathway involving ZO-1 or ZO-1-like protein(s).

Finally, because of their abilities to stimulate elongation by RNA polymerase II through a wide variety of DNA template sequences, TFIIF, Elongin (SIII), and ELL have been considered "general" transcription factors. Our finding that the *ELL2* and *ELL* genes are expressed in many of the same tissues, but that the ratio of ELL2 and ELL mRNAs exhibits tissue to tissue variation, raises the possibility that ELL2 and ELL may perform gene- or tissue-specific functions.